A Multi-Method Exploration of Crime Hot Spots: An Evaluation of the "Repeat Places" Mapping Technique[1]

By
Elizabeth Groff

with
Nancy La Vigne, Cynthia Nahabedian, Eric Jefferis,
Maureen O'Connell, Joseph Szakas, and Julie Wartell

Crime Mapping Research Center
National Institute of Justice
810 7th Street, NW
Washington, DC 20531
202.305.3301
groffe@ojp.usdoj.gov

Presented at the
Annual Meeting of the *Academy of Criminal Justice Sciences (ACJS)*
Albuquerque, New Mexico

March 11, 1998

[1] Points of view are those of the author and do not necessarily represent the view of the US Department of Justice or the National Institute of Justice

INTRODUCTION

The following paper is one of a series that compare various methods for identifying hotspot areas of crime using the same data set. Specifically, the paper evaluates the use of the "repeat place" mapping approach to identify hotspots as developed by Eck, Gersh and Taylor (1997). The "repeat place" technique's basic premise is that crimes are concentrated at specific places (i.e. addresses, intersections, corners, etc.) even within high crime areas. By focusing on those places that account for the highest proportion of crimes and discarding the remainder of crime locations in the data set, the authors believe it is possible to more easily identify both the problem locations and any clustering of those locations.

It is important to note that this is a 'specific place' based approach to identifying hot spots rather than an area based approach like the others that are part of the hotspot evaluation. Maps that depict the results of an analysis using "repeat places" are still point-based, and the only method for identifying clustering is visual inspection by the map user.

The evaluation begins with a detailed explanation of the steps involved in utilizing the "repeat place" mapping technique. Next, the results of the analysis are discussed, and finally, there is a discussion about the utility of this technique.

METHODOLOGY

Before discussing the application of the "repeat place" technique, it is important to understand both the technique itself and the methodology that governed the comparison of hotspot methods. Fittingly, this section begins with a general introduction to the "repeat place" technique. Next, the guidelines for evaluating each method are covered and some background information about the data is detailed. Finally, specifics about the software and scales used in the analysis are described.

Use of the "repeat place" method involves identifying the places in a data set that account for the highest proportion of crime and then plotting only those locations on a map. To identify the highest proportion of crime locations, Eck, Gersh and Taylor (1997) suggest beginning with the top 10% of places, "a convenient number," as a guide. Once these places are identified via a simple visual inspection of the reverse cumulative distribution of places reporting crimes, the next step is to look at the number of events that are occurring across these highest crime places. The goal is to figure out the MPD or "minimum plotting density" that is closest to encompassing 10% of the locations. Eck, Gersh and Taylor (1997) define MPD as the minimum number of events at a place that the analyst needs to consider it a hotspot. Another way to look at the MPD is simply as a way to decide which places will be plotted on the map. For instance, if the least number of incidences at a location within the top 10% is 4 then only those locations with 4 or more crime events will be plotted.

In order to make the evaluation of hotspot methods more meaningful each evaluator is using the same set of data. The data set is for two specific crime types, in one county, over a 13 month period of time. Specifically, the data set includes street robberies and burglaries for Baltimore County, Maryland over the period of November, 1996 to November, 1997. All events were geocoded by Baltimore County Police Department.

Since "repeat place" involves a technique rather than a specific software program, no specialized software, beyond a basic mapping software package and a spreadsheet/statistical program, is required. As such, the "repeat place" method can be easily programmed to work with various types of geographic information systems and desktop mapping programs. In their study, Eck, Gersh and Taylor (1997) utilized MapInfo®. For this evaluation, both ArcView® and EXCEL® are utilized.

In practice, "repeat place" identification is a multi-step process that involves manipulation of the standard data in ArcView, exporting that data to EXCEL, calculating the statistics in EXCEL, and then mapping the identified locations in ArcView. See below for a more detailed explanation.

While all the data are for Baltimore County, four separate analyses are conducted. Both street robberies and burglaries are examined at the county level and then at the sub-county level. The larger scale analysis considers only the events that fall within the southwestern portion of the County just west of Baltimore City.

The most straightforward method of explaining this technique is to run through the steps using an example. This example, which uses the data for burglaries in Baltimore County, will illustrate the process of identifying an MPD for a point distribution.

STEP 1

Take the file of events and aggregate them to the location at which they occurred. The developers of this technique use an address field to do this aggregation. They caution potential users, however, about the inherent pitfall of using addresses. Namely, data entry of addresses is not usually standardized. For example the following addresses: 101 East Fourth Street and 101 E. 4th Street are obviously the same address but would not be recognized as such by a standard aggregation routine in a computer. Consequently, extensive editing must be completed before addresses can be standardized well enough for aggregation on the address field to output reliable results. Because of the potential problem with aggregating on the address field, this author used the x and y coordinates of each point to create a unique value for each one of the geocoded points. The basis for this unique value was the concatenation of the x_coord and the y_coord items. This concatenated item was then the case item on which the crime reports were aggregated (using the Sum tool in ArcView). The resulting table, which contains counts of crime events at each xy location, is then ready to be opened in EXCEL.

STEP 2

Open the summary table in EXCEL. The table has two fields: XY_Coord and Count (Table 1). Sort the table on the Count field in descending order. The locations with the greatest number of crimes should now be at the top of the column. The next step is to calculate a new field (Table 2: Column C) that contains the proportion of events that happened at each location (Formula 1). This figure provides the analyst with a quick reference to the proportion of crimes accounted for by any specific location in the distribution.

Table 1: Count Table of Burglary Incidents at a Specific Location

A XY_Coord	**B** Count of Burglary Crime Events
764512393558	16
765971394100	16
767293393432	16
764710393741	9
766015394043	9
763971393374	8

Formula 1

Proportion of Events = Number of events at a location / Total number of events in file
Or
Column C = Column B / 6219

Table 2: Burglary Analysis Table

A XY_Coord	**B** Number of Crime Events	**C** Proportion of Events	**D** Cum Total of Events	**E** Reverse Cumulative Distribution	**F** Cumulative Total of Places	**G** Cumulative Distribution of Places	**H** Mapping Efficiency
764512393558	16	0.0026	16	0.0026	1	0.0002	0.9262
765971394100	16	0.0026	32	0.0051	2	0.0004	0.9262
767293393432	16	0.0026	48	0.0077	3	0.0006	0.9262
764710393741	9	0.0014	57	0.0092	4	0.0008	0.9172
766015394043	9	0.0014	66	0.0106	5	0.0009	0.9106
763971393374	8	0.0013	74	0.0119	6	0.0011	0.9043
766020393906	8	0.0013	82	0.0132	7	0.0013	0.8993
766535397143	8	0.0013	90	0.0145	8	0.0015	0.8951
766605392373	8	0.0013	98	0.0158	9	0.0017	0.8916
764785393500	7	0.0011	105	0.0169	10	0.0019	0.8876

STEP 3

Calculate another field to represent the cumulative total of events (Formula 2). This field provides an easy reference for statements about how many events accounted for a certain proportion of total crime (Table 2: Column D).

Formula 2

Cumulative Total of Events = Number of events at a location + Number of events at the next location

STEP 4

In Column E, calculate a reverse cumulative distribution of events. Basically, this keeps a running proportion of events so that the reader can tell what percentage of all crime events have been accounted for at any point in the distribution (Table 2). There were 6,219 burglaries reported and successfully geocoded during the study period.

Formula 3

Reverse Cumulative Distribution of Events =

Reverse cumulative total of events / Total number of events in file

Or

Column E = Column D / 6219

STEP 5

At this point, the focus moves from the distribution of events to the distribution of locations. Column F contains the cumulative number of locations/places in the file (Table 2). This is easily accomplished with the series function in EXCEL.

STEP 6

Now, the analyst will need to compute the cumulative distribution of places and place the result in Column G (Formula 4). In other words, this field tells the user what percentage of places have been accounted for at any point in the distribution (Table 2). There were 5,270 unique places reporting burglaries during the study period.

Formula 4

Cumulative Distribution of Places = Cumulative total of places / Total number of places

Or

Column G = Column F / 5270

STEP 7

Finally, a measure of "mapping efficiency" is calculated that describes the degree to which the resulting map depicts the distribution of places to events. The closer the calculated mapping efficiency score is to 1, the fewer points are plotted on the map. Those few points represent a large proportion of the events (in this case burglaries). Consequently, the map has higher utility to the end user (Eck, Gersh and Taylor, 1997). The mapping efficiency value is calculated by subtracting one from the ratio of places to crimes (Formula 5; Table 2: Column H).

Formula 5

Mapping Efficiency = 1**-** (Cumulative distribution of places / Reverse cumulative distribution of events)
Or
Column H = 1**-** (Column G / Column E)

STEP 8

At this point the analyst must examine the distribution and identify the top 10% of places (i.e. the 10% of places with the highest number of crime events). This is accomplished by looking for the 10% value in the cumulative distribution of places field (Column G). Table 3 shows the portion of the table at which the cumulative places account for approximately 11.7% to 11.9% of all places at which burglaries were reported (Table 3: Column F). The record at which the cumulative number of places account for 10% of the distribution is actually 'above' the section shown in Table 3.

Once this 10% record is identified the analyst then checks the total number of events at that place (Table 3: Column B). In this example, at the "place" where the distribution hit 10%, a total of 280 events had occurred across 105 locations (portion of the table not shown). The number of crimes reported at that specific location was 2. However, the ultimate goal of "repeat place" technique is to identify the MPD for a distribution. So, if the 10% guide falls in the middle of a series of locations with the same number of events, the analyst must then look for the next change in the number of incidents at a place. It is the analyst's decision whether they look "up" (i.e. more events at a place) or "down" " (i.e. fewer events at a place) in the distribution to see at what point the number of incidents changes (Table 3: shaded rows). Once this point of change is identified, the number of events at that location becomes the MPD. In this example of burglaries, an MPD of 2 was identified. This encompassed the 11.8% (622) of places that accounted for 25.3% (1571) of events.

The mapping efficiency of .53 at this number of incidents was fairly low (Map 1). Once again, the higher the mapping efficiency, the greater the proportion of events that are represented by points that are plotted on the map. The goal is to plot the fewest possible locations that represent the highest proportion of burglaries.

Table 3: Change Point in Distribution of Number of Crime Events Per Location

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
XY_Coord	Number of Crime Events	Proportion of Events	Cum Total of Events	Reverse Cumulative Distribution	Cum total of Places	Cumulative Distribution of Places	Mapping Efficiency
768320394688	2	0.0003	1565	0.2516	619	0.1175	0.5332
768424395201	2	0.0003	1567	0.2520	620	0.1176	0.5331
768489394421	2	0.0003	1569	0.2523	621	0.1178	0.5329
768619393352	2	0.0003	1571	0.2526	622	0.1180	0.5328
763404393705	1	0.0002	1572	0.2528	623	0.1182	0.5323
763405393704	1	0.0002	1573	0.2529	624	0.1184	0.5319
763453393588	1	0.0002	1574	0.2531	625	0.1186	0.5314

763522393757	1	0.0002	1575	0.2533	626	0.1188	0.5310

RESULTS

This section covers the outcome of conducting a "repeat place" analysis on street robberies and burglaries at two different geographic levels. Burglary results are discussed first and then street robbery.

Burglary

The results of the burglary analysis are summarized in Table 4. At the County level, there were 6,219 burglaries that were reported at 5,280 separate locations (Map 1). The smaller geographic area had 2,009 burglaries occurring at 1,722 places (Map 2). A visual inspection of these two sets of numbers clearly indicates that events are dispersed over many places. The analysis proved this deduction. Because of the low proportion of locations with three or more events, the top ten percent included many locations with two events. This meant that many locations (622 for the County and 201 for Southwest) were plotted on the map (Table 4; Maps 1 and 2). It is interesting that the change in scale did not significantly change the proportion of events encompassed in the top 10% places nor did it change the MPD or Mapping Efficiency measures. A visual inspection also revealed no significant changes in pattern between the two scales (Maps 1 and 2). However, a look at the legend clearly shows that the hottest locations for burglaries are not in Southwest Baltimore County (see maximum hot places in County legend [N=240] compared to sub-county legend [N=16]).

Table 4: Results of Burglary Analysis

	Events		**Places**		**MPD**	**Mapping Efficiency**
Geographic Level	**Proportion**	**Number**	**Proportion**	**Number**		
County	25%	1571	12%	622	2+	.53
Southwestern Baltimore County	24%	488	12%	201	2+	.52

Robbery

Table 5 contains the results for street robberies. Total street robberies for the study period was 1,223 at 1,048 places. In the southwestern area, there were 485 robberies at 415 locations. Once again, there was only a small proportion of locations with more than two events. The proportions of places to crimes at the two scales were identical as was the MPD. The mapping efficiency varied only slightly between the two scales. As with the burglary results, there were no significant changes in spatial pattern but the hottest places were eliminated when only southwest Baltimore County was considered (Maps 3 and 4).

Table 5: Results of Robbery Analysis

	Events		**Places**		**MPD**	**Mapping Efficiency**
Geographic Level	**Proportion**	**Number**	**Proportion**	**Number**		
County	23%	277	10%	102	2+	.57
Southwestern Baltimore County	23%	113	10%	43	2+	.56

CONCLUSIONS

The following section provides a discussion of "repeat place" mapping in the context of this study. Topics covered include ease of use, validity of results, practical utility and flexibility of the technique and overall conclusions.

Ease of Use

This method is very straightforward and is easy to automate. While the steps are not difficult to understand, it is time consuming to proceed through the many steps (i.e. add fields, concatenate items, export files, open another software program, enter formulas, identify the appropriate value and then go back to the mapping program). However, if automated, the speed of the procedure would be limited only by factors external to the methodology (e.g. the processing speed of the machine, network speed or size of files being analyzed). File types used in the technique are standard and read easily into a variety of software programs.

Validity of Results

The technique is clearly helpful at identifying problem "places" in an area. There are no complex formulas behind the method. This simplicity makes it easy to check the results against the actual distribution of crime events. The locations identified as problem places were valid ones. One aspect to note is that the method does not take into account the number of possible targets at an address. Hence, the method may primarily be capturing places with a greater population density (e.g. multiple units in apartment buildings or public housing) rather than those that have an unusually high rate of victimization.

Overall, the method is quantifiable and objective. A significant degree of subjectivity in introduced into the method when choosing a minimum plotting density (MPD), which effects the visual representation and the information conveyed by the analysis. Thus choosing a higher plotting density will emphasize those locations with the highest number of crimes. However, a high plotting density also removes all information about locations with MPD's below that threshold. This information is effectively lost since it is not plotted on the map for the end user's consideration. Loss of information is especially problematic when considering how those unplotted locations may be instrumental in identifying 'hot blocks' in addition to the 'repeat places' that were plotted.

Practical Utility

The sticky part of including this method in a comparison of methods for identifying "hotspots" of crime is that most of the "hotspots" identified by other software packages are actually "hotspot areas." This method represents more of a repeat incident analysis. In other words, it identifies specific addresses at which there were many crimes reported. There are two main challenges in using this method; type of distribution being analyzed and lack of an objective determination of clusters. The method seems to work best with distributions in which a few places account for many incidents. If an analyst is trying to use this technique on a large distribution in which many places are victimized a few times and a relatively few number of places are victimized many times, there may be so many 'repeat places' identified that the resulting map is still very cluttered. Second, there is no objective method included in the technique with which the analysis can identify clusters of repeat places. Thus the technique excels at identifying specific places but not at objectively identifying problem areas.

Flexibility of the Technique

Because this is a technique and not a canned program it is easily transferable to an infinite number of software products and product configurations. All that is required for implementation of this method is to follow the steps that are clearly outlined by the authors (Eck, Gersh and Taylor, 1997).

Summary

This technique could have great utility in policing where often the goal is to attack specific places that are perennial problems. It is excellent at focusing attention on the most frequent locations of activity. This is especially helpful when the distribution being analyzed has a few locations with many occurrences.

However, as a general technique for identifying areas of concentrated crime, the method should be used with some caveats in mind. First, the method does not identify 'hot areas' it identifies 'repeat places'. Second, the elimination of data points that do not meet the MPD can be positive or negative depending on the type of event being examined. Displaying only the points above MPD does cut down on the visual clutter of a map. On the other hand, the loss of information makes it difficult to identify where there are visual clusters of incidents. In distributions where the location of events relative to one another is less important than the frequency with which the event occurs at the same location, this method is very strong. While in distributions where locations of events are related to one another, the elimination of locations from the analysis makes it difficult to identify areas of concentration. Often in crime patterns, it is not simply the one place with ten incidents that is indicative of a crime problem. Rather it is the aggregate picture of the one ‘repeat place’ in addition to ten other places with one crime each on the same block face that is critical. This method would eliminate those places with one event since it gives no consideration to the relative positioning of places. This is especially regrettable given the holistic nature of crime. Events at neighboring places usually contribute to the problem at a specific place and it is often the aggregate that best describes a crime problem. The third caveat concerns the subjectivity inherent in the determination of the MPD. The point at which the MPD cutoff is established greatly influences the resulting map and it is largely governed by the analyst.

In sum, this technique seems more suitable to examining crime types or events that tend to concentrate in a few places rather than dispersed ones. Unfortunately, the crimes chosen for this evaluation were very dispersed so the technique performed only marginally well.